NORMAL CLASSIFICATION OF 3D OCCUPANCY GRIDS FOR VOXEL-BASED INDOOR RECONSTRUCTION FROM POINT CLOUDS

P. Hübner, S. Wursthorn, M. Weinmann

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

1. INTRODUCTION

In recent years, digital models of indoor building environments (Bormann et al., 2018) have experienced an ever increasing surge in importance (Chaffamahoseini et al., 2017; Sacks et al., 2020) in different fields of application, such as construction (Jafari et al., 2021), facility management (Gao and Pishdad-Bozorgi, 2019), energy efficiency (Jin et al., 2019) or cultural heritage (Solla et al., 2020). In this context, methods for the efficient creation of building models for existing building structures (Volk et al., 2014) have come to the focus of current research efforts (Lehtola et al., 2020; Weinmann et al., 2021) in the fields of indoor mapping (Otero et al., 2020) (i.e. the efficient acquisition of 3D indoor building geometry by means of mobile sensor systems) and indoor reconstruction (Jiang et al., 2020; Pintore et al., 2020) (i.e. the automated generation of building models from indoor mapping data).

Information hinting on a distinction between interior and exterior with respect to building surfaces represented in indoor mapping geometry can offer a valuable guidance for automated indoor reconstruction approaches. This information can be provided by normal vector directions, if they are consistently oriented with respect to the building interior. While normal vectors can be efficiently determined for points of a point cloud by analyzing the point distribution in the neighbourhood of each respective point (Yu et al., 2019; Sanchez et al., 2020a), consistently determining the absolute orientation (i.e. pointing forwards or backwards along the determined direction) can be more challenging (Ochmann and Klein, 2019).

When an indoor mapping system provides information about the position of the sensor at the time of recording of each respective point, the normal vector can be oriented towards the sensor. Often, however, only the point coordinates are provided. Other indoor mapping systems provide output in the form of triangle meshes as derived product generated in a black-box process from the primary sensor measurements like e.g. the Microsoft HoloLens (Khoshelham et al., 2019; Hübner et al., 2020a) or the Matterport system (Chang et al., 2017). These triangle meshes were found to be comparable with point clouds in regard of their aptitude towards classification and segmentation tasks while representing a significantly more compact form of data (Bassier et al., 2020; Weinmann et al., 2020).

Furthermore, these triangle meshes provide consistently oriented normal vectors. A recently published voxel-based indoor reconstruction approach (VoxIR) (Hübner et al., 2020b, 2021a) with publicly available code is tailored towards such indoor mapping triangle meshes and depends on their normal vectors as input data. However, it is not straightforward to apply this approach to indoor mapping point clouds, where consistently oriented normal vectors are not available.

Thus, we provide in this paper the following contributions:

- We present a novel method for the automated classification of 3D occupancy voxel grids of indoor building environments according to their normal direction in vertical and horizontal structures. Voxels with vertical normal direction are further distinguished in those, whose normal direction is directed upwards, downwards or both with respect to the building interior.
- We apply the proposed normal classification method in the context of voxel-based indoor reconstruction, extending an established approach so far only applicable to triangle meshes with normal vectors to indoor mapping point clouds without normal information.
- We present a thorough qualitative and quantitative evaluation on two different publicly available benchmark datasets for indoor reconstruction.

In the following, Section 2 gives a brief overview on related work. Afterwards, the proposed methodology is explained in Section 3 before Section 4 presents both qualitative and quantitative evaluation results which are further discussed in Section 5. Finally,

KEYWORDS: Indoor Reconstruction, Voxel, Building Model, Normal Vector, Point Cloud, Triangle Mesh

ABSTRACT:

In this paper, we present an automated method for classification of binary voxel occupancy grids of discretized indoor mapping data such as point clouds or triangle meshes according to normal vector directions. Filled voxels get assigned normal class labels distinguishing between horizontal and vertical building structures. The horizontal building structures are further differentiated into those with normal directions pointing upwards or downwards with respect to the building interior. The derived normal grids can be deployed in the context of existing voxel-based indoor reconstruction pipeline, which so far was only applicable to indoor mapping triangle meshes that already contain normal vectors consistently oriented with respect to the building interior. By means of quantitative evaluation against reference data, we demonstrate the performance of the proposed method and its applicability in the context of voxel-based indoor reconstruction from indoor mapping point clouds without normal vectors. The code of our implementation is made available to the public at https://github.com/huepat/voxir.

1. INTRODUCTION

1. INTRODUCTION

In recent years, digital models of indoor building environments (Bormann et al., 2018) have experienced an ever increasing surge in importance (Chaffamahoseini et al., 2017; Sacks et al., 2020) in different fields of application, such as construction (Jafari et al., 2021), facility management (Gao and Pishdad-Bozorgi, 2019), energy efficiency (Jin et al., 2019) or cultural heritage (Solla et al., 2020). In this context, methods for the efficient creation of building models for existing building structures (Volk et al., 2014) have come to the focus of current research efforts (Lehtola et al., 2020; Weinmann et al., 2021) in the fields of indoor mapping (Otero et al., 2020) (i.e. the efficient acquisition of 3D indoor building geometry by means of mobile sensor systems) and indoor reconstruction (Jiang et al., 2020; Pintore et al., 2020) (i.e. the automated generation of building models from indoor mapping data).

Information hinting on a distinction between interior and exterior with respect to building surfaces represented in indoor mapping geometry can offer a valuable guidance for automated indoor reconstruction approaches. This information can be provided by normal vector directions, if they are consistently oriented with respect to the building interior. While normal vectors can be efficiently determined for points of a point cloud by analyzing the point distribution in the neighbourhood of each respective point (Yu et al., 2019; Sanchez et al., 2020a), consistently determining the absolute orientation (i.e. pointing forwards or backwards along the determined direction) can be more challenging (Ochmann and Klein, 2019).

When an indoor mapping system provides information about the position of the sensor at the time of recording of each respective point, the normal vector can be oriented towards the sensor. Often, however, only the point coordinates are provided. Other indoor mapping systems provide output in the form of triangle meshes as derived product generated in a black-box process from the primary sensor measurements like e.g. the Microsoft HoloLens (Khoshelham et al., 2019; Hübner et al., 2020a) or the Matterport system (Chang et al., 2017). These triangle meshes were found to be comparable with point clouds in regard of their aptitude towards classification and segmentation tasks while representing a significantly more compact form of data (Bassier et al., 2020; Weinmann et al., 2020).

Furthermore, these triangle meshes provide consistently oriented normal vectors. A recently published voxel-based indoor reconstruction approach (VoxIR) (Hübner et al., 2020b, 2021a) with publicly available code is tailored towards such indoor mapping triangle meshes and depends on their normal vectors as input data. However, it is not straightforward to apply this approach to indoor mapping point clouds, where consistently oriented normal vectors are not available.

Thus, we provide in this paper the following contributions:

- We present a novel method for the automated classification of 3D occupancy voxel grids of indoor building environments according to their normal direction in vertical and horizontal structures. Voxels with vertical normal direction are further distinguished in those, whose normal direction is directed upwards, downwards or both with respect to the building interior.
- We apply the proposed normal classification method in the context of voxel-based indoor reconstruction, extending an established approach so far only applicable to triangle meshes with normal vectors to indoor mapping point clouds without normal information.
- We present a thorough qualitative and quantitative evaluation on two different publicly available benchmark datasets for indoor reconstruction.

In the following, Section 2 gives a brief overview on related work. Afterwards, the proposed methodology is explained in Section 3 before Section 4 presents both qualitative and quantitative evaluation results which are further discussed in Section 5. Finally,
the paper closes with a concluding summary and suggestions for future research in Section 6.

2. RELATED WORK

Automated reconstruction of building environments from indoor mapping data such as point clouds is a wide and active field of research (Kang et al. 2020; Pintore et al. 2020). The various proposed approaches differ significantly in the amount of assumptions that are made with respect to the building structure to be reconstructed and thus in their flexibility towards challenging building environments, ranging from single room scenarios (Li et al. 2020; Sanchez et al. 2020b). Manhattan World structures where all surfaces are orthogonal to the coordinate axes (Ryu et al. 2020; Kim et al. 2020) to diagonal (Shi et al. 2019; Tran and Khoshelham 2020) or even curved walls (Yang et al. 2020) and slanted ceilings (Nikoohemat et al. 2020; Lim and Doh 2021).

The available methods also differ in their general approaches. Some follow a bottom-up strategy, where small plane patches are detected in the point cloud data and assembled to constitute rooms (Xie et al. 2019; Shi et al. 2020; Oh et al. 2021). Others follow a more top-down approach by first detecting the dominant global planes in the dataset and intersecting them with one another. The cells of the resulting cell complex are then partitioned into building interior and outside space. Storey-wise 2D cell complexes (Li et al. 2018; Tran and Khoshelham 2020) as well as fully three-dimensional cell complexes can be used (Coudron et al. 2018; Ochmann et al. 2019). Other reconstruction methods make use of trajectory information of the mobile mapping system if available (Kui et al. 2019; Nikoohemat et al. 2020; Lim and Doh 2021) or operate in a discretized voxel grid (Fichtner et al. 2017; Flikweert et al. 2019; Gorte et al. 2019). Recently, reconstruction methods relying on deep learning methods are gaining prevalence (Lim et al. 2020; Gankhuyag and Han 2021; Yang et al. 2021).

Regarding the task of determining normal directions with a consistent orientation with respect to the building interior as a preprocessing step to indoor reconstruction (i.e. not deriving the information about the normal orientation from the reconstructed indoor models), only few works are that are dedicated specifically to this topic (Ochmann and Klein 2019). The task of determining floor and ceiling layers from indoor mapping point clouds, on the other hand, is addressed more frequently in the context of indoor reconstruction (Macher et al. 2017; Fichtner et al. 2017; Elsegy et al. 2018; Li et al. 2018; Lemo et al. 2019; Romero-Jarén and Arranz 2021). Usually, however, the ceilings and floors are assumed to occur in structure of fixed storey levels globally over the whole building and are often assumed to be planar.

3. METHOD

In the following, we present a novel methodology to classify the filled voxels of 3D occupancy grids derived from indoor mapping data according to their main normal direction with respect to the represented building structures. First, Section 3.1 describes the proposed approach. Afterwards, Section 3.2 describes its application in the context of voxel-based indoor reconstruction.

3.1 Normal Classification of Occupancy Grids

The aim of the approach presented here is to classify the filled voxels of a 3D occupancy grid representing indoor building environments into vertical structures (of horizontal normal direction class $N_H$) and horizontal structures (of vertical normal direction $N_V$). The $N_V$ voxels are further subdivided into the normal classes $N_U$ (normal up), $N_D$ (normal down) and $N_{UD}$ (normal up and down) with regard to their normal direction pointing towards the room interior. Here, the normal class $N_{UD}$ considers voxels vertically subdividing rooms of different storeys, i.e. where a ceiling surface and the next floor surface above are covered by one and the same voxel.

As is the case with the indoor reconstruction procedure VoxIR (Hübner et al. 2020b, 2021a) in which the proposed normal classification method is to be deployed, we do not assume room surfaces to be planar or subject to common restrictions like the Manhattan World assumption (Hübner et al. 2021b). In particular, we do not assume that room surfaces are necessarily structured in distinct storeys but account for floors and ceilings to be on different height levels over different rooms as well as within one room.

The proposed normal classification approach is further detailed in the following sections. A graphical overview is presented in Figure 1, while exemplary results for a section of the dataset ‘Office’ of Hübner et al. 2021a are depicted in Figure 2.

3.1.1 Vertical Structure Extraction In a first step, clearly vertical structures such as wall surfaces are detected in the input voxel occupancy grid and labeled as $N_Y$. To this aim, continuous vertical columns of filled voxels of at least 0.5 m of height are detected. This parameter choice was established as producing satisfactory results over a range of different datasets. Exemplary results are depicted in Figure 2(a).

3.1.2 Horizontal Structure Segmentation The remaining voxels not labeled as $N_Y$ are preliminarily assumed to belong to horizontal structures (i.e. $N_V$) and are segmented into horizontal voxel segments $S_V$ depicted in Figure 2(b). As in VoxIR, a 2.5D
region growing with a threshold on height differences between positionally neighbouring voxels is used, in order to allow $S_v$ to stretch over small height offsets such as stair steps. This can however lead to $S_v$ under-segmenting ceiling surfaces together with the floor surface of the next room above. To prevent this, an additional stopping condition is introduced for the region growing process by not only regarding height differences in the voxel segment itself but also differences in the vertical distance above the respective voxels until the next voxels not yet classified as $N_H$ (or the upper border of the voxel grid) are encountered.

3.1.3 Vertical Adjacency Graph The resulting horizontal segments $S_V$ are assembled in a Vertical Adjacency Graph (VAG) where the voxel segments are the nodes and the edges represent a vertical adjacency (i.e. above/below) relation between them, weighted with the respective area of coverage and mean vertical distance. Besides the $S_V$, the upper and lower outside $O_D$ and $O_U$ are included as two additional nodes in the graph as neighbour segments that to not have another segment above/below them. The resulting graph serves as input to the subsequent iterative step of normal classification.

3.1.4 Iterative Normal Classification of Horizontal Segments Some of the $S_V$ as nodes of the VAG are assigned normal classes $N_U$, $N_D$ and $N_{UD}$ in an iterative process aiming at identifying main layers of segments $S_U$ and $S_D$ which are clearly definable as deliminating rooms vertically from below and above, respectively. Weighted these layers pass vertically upwards through the voxel grid during the iterative process, they are not necessarily restricted to a single height level at a time (i.e. storey-wise).

The iterative process starts from below as we assume the lower surfaces of a room (i.e. floor and horizontal furniture surfaces acquired from above by the respective indoor mapping system) to be more complete than the upper surfaces. The process can however be inverted starting from above if this is more suited for a given indoor mapping system. All examples and results presented in this work are processed from bottom to top.

First, the first lower layer $S_{U_0}$ is initialized by detecting $S_V$ whose largest neighbour below with respect to coverage is the lower outside node $O_U$, of the VAG. Starting from here, each lower layer $S_{U_i}$ is classified as $N_U$, and a corresponding upper layer $S_{D_i}$ is determined as those $S_V$ whose largest neighbour below is one of the $S_{U_i}$ segments. Here, however, a minimum threshold of 1.5 m on the mean vertical distance between $S_V$ is applied, as we associate $S_{U_i}$ and $S_{D_i}$ with lower and upper surfaces of rooms and assume rooms to have a certain minimum height. The detected $S_{D_i}$ are assigned the $N_D$ label.

In each iteration, further $S_V$ between the detected main layer segments can be classified as $N_U$ or $N_D$ if its largest upper neighbour is among the current $S_{D_i}$ and its largest lower neighbour is among the $S_{U_i}$ and is also a lower neighbour of the respective upper $S_{D_i}$ neighbour. For the classification of an intermediate segment detected in this manner, its lateral borders are examined in the height range between its $S_{U_i}$ and $S_{D_i}$ neighbours. If there are voxels indicating a lateral vertical surface below/above the $S_i$ segment, it is assumed to represent the upper/lower surface of a piece of furniture (e.g. table surface / lower surface of a lamp) and assigned $N_U$ or $N_D$ accordingly. If there are no lateral surfaces to be found, the decision is made based upon height above the lower $S_{U_i}$ neighbour, assuming that horizontal surfaces below 1.5 m above the floor are likely to be captured from above and thus assigned $N_U$.

Based on the current $S_{D_i}$, the lower main layer of the next iteration $S_{U_{i+1}}$ is detected, again as those $S_V$, whose largest lower neighbour is among the $S_{D_i}$. Here, however, it needs to be considered that depending on voxel resolution and width of building structures, ceiling surfaces and the next floor surfaces above could be covered by the same horizontal layer of voxels. Thus, if the mean vertical distance from a $S_{D_i}$ segment to its $S_{U_{i+1}}$ neighbour above is more than 1.5 m, the respective $S_{D_i}$ segment gets assigned an additional $N_{UD}$ label and is itself part of $S_{U_{i+1}}$ instead of its upper neighbour.

This process terminates, when no $S_{U_{i+1}}$ can be found anymore. Intermediate results of the iterations are exemplarily depicted in Figure 2(c) to (f).

3.1.5 Completion and Refinement $S_V$ segments that are so far not yet assigned a normal class value (e.g. because due to occlusion or incomplete acquisition of building geometry they...
do not have consistent upper and lower neighbours among the main layers of the same iteration) are finally considered in this step. All direct voxel contacts of an unclassified segment with other voxels that are already assigned a value \(N_{V_I}\) as well as \(N_{V_T}\) are considered and counted. A segment is assigned to the class with the largest contact count. If it does not directly contact any classified voxels, it is assigned \(N_{V_T}\). The final classification result of the \(S_{V_I}\) segments is depicted in Figure 2(a).

Lastly some refinement steps are conducted on the resulting normal grid. For instance, vertical pillars of multiple \(N_{V_I}\) voxels are resolved by leaving only the topmost (for \(N_{V_T}\)) or lowest (for \(N_{V_T}\)) voxel \(N_{V_T}\) and turning the others to \(N_{V_T}\) as they clearly form a vertical structure. The final normal grid is exemplarily depicted in Figure 2(h).

### 3.2 Voxel-based Indoor Reconstruction from Point Clouds without Normal Vectors

The voxel-based indoor reconstruction method VoxIR presented in (Hübner et al., 2020a, 2021a) is tailored towards indoor mapping triangle meshes as acquired for instance with the Microsoft HoloLens (Hübner et al., 2020a) or the Matterport system (Chang et al., 2017) as these provide normal vectors consistently oriented with respect to inside/outside of the building structure. These oriented normals are further discretized to a normal grid of arbitrary resolution with the values \(N_V\), \(N_{H}\), and \(N_D\) as described above which is the input to the voxel-based indoor reconstruction process.

In the scope of this work, we extend VoxIR to also be applicable to point clouds without normal vectors by means of the normal classification approach on occupancy grids presented in Section 3.1. Furthermore, we extend VoxIR to consider \(N_{H}\)/\(D\) voxel representing a ceiling surface as well as the next floor above. The code of the extended version of VoxIR adapted to indoor mapping point clouds such as those of the ISPRS Benchmark on Indoor Modeling (Khoshelham et al., 2017) including our implementation of the normal classification procedure described above is released at https://github.com/huepat/voxir

### 4. EVALUATION AND RESULTS

In order to quantitatively evaluate our proposed approach, we use the metrics of completeness, correctness and accuracy as proposed in (Khoshelham et al., 2018) and used in the context of evaluating the contributions to the ISPRS Challenge on Indoor Modeling (Khoshelham et al., 2020a, 2021a). However, these metrics expect reference and reconstructed building geometry to be given in continuous, Euclidean space, considering the fraction of reconstructed building structures within a given buffer distance around the reference building structure. As our method operates in discrete voxel space, we also use discretized versions of the evaluation metrics.

Completeness \(b\) is calculated as:

\[
\text{Completeness}(b) = \frac{\#V_{\text{Rec}(b)}}{\#V_{\text{Ref}}} \tag{1}
\]

Accuracy \(b\) is calculated as:

\[
\text{Correctness}(b) = \frac{\#V_{\text{Rec}(b)}}{\#V_{\text{Ref}}} \tag{2}
\]

with \(\#V_{\text{Rec}}\) denoting the number of voxels representing building geometry in the discretized reference, \(\#V_{\text{Rec}}\) the same value for the reconstruction voxel grid and \(\#V_{\text{Rec}}\) being the number of true positive reconstructed voxels with respect to a buffer level \(b\). For \(b = 1\), we only consider a voxel as a true positive, if the corresponding voxel position in the reference grid reports building structure as well. For \(b > 1\), we also consider voxels as true positives, where reference building structure is found within a \((2b)\)-neighbourhood of the respective voxel position. For \(b = 2\), we consider the direct neighbours of the voxel position and for \(b = 3\) also the neighbours of these neighbours, and so on.

Furthermore, we present accuracy values in analogy with (Khoshelham et al., 2018), quantifying the distance \(d\) of reconstructed building geometry to corresponding reference geometry depending on the buffer level \(b\):

\[
\text{Accuracy}(b) = \text{Mean}(d(b)) \tag{3}
\]

In our case, we use as distance \(d\) the distance between the respective voxel center coordinates. Furthermore, instead of using the median, we chose to use the arithmetic mean, as possible distances in our case are restricted to a few possible values due to the discretization of the voxel grid.

While in (Hübner et al., 2021a) quantitative evaluation results are presented for different semantic classes, here, we consider only the building structure without any further semantic distinction as proposed in (Khoshelham et al., 2018). In doing so, we consider the semantic classes of ‘Wall’, ‘Floor’ and ‘Ceiling’ as belonging to the building structure, while ‘Interior Object’ representing furniture and clutter is disregarded as well as ‘Empty Interior’ and ‘Wall Opening’. For all presented experiments, a fixed voxel resolution of 5 cm is used. Investigating the impact of this parameter exceeds the scope of this work and is left to future research.

### 4.1 Normal Classification of Occupancy Grids

First, we evaluate completeness and correctness of the normal classification procedure presented in Section 3.1 on indoor mapping triangle meshes, where normal vectors and reference values are available. To this aim, we use the four publicly available datasets of HoloLens triangle meshes of different indoor environments presented in (Hübner et al., 2021a), where further details on the datasets can be found. The respective datasets are discretized to binary occupancy grids and, for each filled voxel, the normal class is determined. As reference data, the voxel grid with normal classification as derived by VoxIR from the triangle meshes and their normal vectors is used. The results are presented in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Completeness</th>
<th>Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>Attic</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>Basement</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>Residential House</td>
<td>0.82</td>
<td>0.80</td>
</tr>
</tbody>
</table>

### 4.2 Impact on VoxIR

The four datasets of triangle meshes used in Section 4.1 are furthermore used to investigate the impact of determining the normal classification from binary occupancy grids instead of from the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.https://doi.org/10.5194/isprs-annals-V-4-2022-121-2022 | © Author(s) 2022. CC BY 4.0 License. 124
This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-4-2022-121-2022 | © Author(s) 2022. CC BY 4.0 License.

Table 2. Evaluation of the impact of the normal grid on VoxIR reconstruction results on the HoloLens triangle meshes published in Hubner et al. [2021a] with 5 cm voxel resolution. The presented values refer to results achieved when using normal grids determined with the method presented in Section 3.1. The values in parentheses refer to results achieved with normal grids determined directly from the triangle meshes.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Buffer Level</th>
<th>Completeness</th>
<th>Correctness</th>
<th>Accuracy [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>1</td>
<td>0.86 (0.87)</td>
<td>0.69 (0.75)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.13 (1.11)</td>
<td>0.91 (0.95)</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.19 (1.14)</td>
<td>0.96 (0.98)</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>Attic</td>
<td>1</td>
<td>0.80 (0.81)</td>
<td>0.65 (0.70)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.05 (1.04)</td>
<td>0.86 (0.90)</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.11 (1.06)</td>
<td>0.91 (0.94)</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>Basement</td>
<td>1</td>
<td>0.83 (0.87)</td>
<td>0.71 (0.72)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.07 (1.11)</td>
<td>0.91 (0.92)</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.12 (1.16)</td>
<td>0.95 (0.96)</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>Residential</td>
<td>1</td>
<td>0.72 (0.83)</td>
<td>0.58 (0.71)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>House</td>
<td>2</td>
<td>0.97 (1.05)</td>
<td>0.78 (0.90)</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.04 (1.09)</td>
<td>0.84 (0.94)</td>
<td>0.03 (0.02)</td>
</tr>
</tbody>
</table>

Table 3. Evaluation of VoxIR reconstruction results on the point clouds published in Khoshelham et al. [2017, 2020] with 5 cm voxel resolution.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Buffer Level</th>
<th>Completeness</th>
<th>Correctness</th>
<th>Accuracy [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>1</td>
<td>0.65</td>
<td>0.34</td>
<td>0.00</td>
</tr>
<tr>
<td>Study</td>
<td>2</td>
<td>1.45</td>
<td>0.77</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.67</td>
<td>0.88</td>
<td>0.06</td>
</tr>
<tr>
<td>Study</td>
<td>1</td>
<td>0.37</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.02</td>
<td>0.63</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.32</td>
<td>0.82</td>
<td>0.08</td>
</tr>
<tr>
<td>Case</td>
<td>1</td>
<td>0.46</td>
<td>0.42</td>
<td>0.00</td>
</tr>
<tr>
<td>Study</td>
<td>2</td>
<td>0.82</td>
<td>0.74</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.89</td>
<td>0.81</td>
<td>0.05</td>
</tr>
<tr>
<td>Case</td>
<td>1</td>
<td>0.43</td>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>Study</td>
<td>2</td>
<td>1.10</td>
<td>0.69</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.34</td>
<td>0.84</td>
<td>0.07</td>
</tr>
<tr>
<td>Case</td>
<td>1</td>
<td>0.49</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>Study</td>
<td>2</td>
<td>1.14</td>
<td>0.71</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.30</td>
<td>0.80</td>
<td>0.06</td>
</tr>
<tr>
<td>Case</td>
<td>1</td>
<td>0.32</td>
<td>0.21</td>
<td>0.00</td>
</tr>
<tr>
<td>Study</td>
<td>2</td>
<td>0.84</td>
<td>0.56</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.11</td>
<td>0.74</td>
<td>0.08</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The results of the evaluation of the normal classification procedure presented in Table 1 hint on a good performance, and Table 2 shows that using these normal grids determined from occupancy grids does not significantly reduce reconstruction quality when fed as input into the VoxIR pipeline. Still, the reconstruction results on the ISPRS benchmark presented in Table 3 leave room for improvement, especially when regarding only the first buffer level, i.e. only considering exact same voxel positions in result and reference grid. However, the quality rapidly improves when considering higher buffer levels.

Overall, the results presented in the right part of Figure 3 show that VoxIR is quite generic with respect to its assumptions on building structures, partly scanned rooms are reconstructed close to the shape of the parts included in the input data. Better reconstruction results are possible when applying more restricting assumptions on building geometry. For instance, the mentioned room in Figure 3(c) could potentially be restored better, when assuming that rooms are rectangular.

On the other hand, this flexibility of VoxIR towards possible building structures can also be regarded as its strength. For instance, the protrusions of the ceiling of the large room in Figure 3(c) are reconstructed quite well, as are the unusual room shapes in Figure 3(f). Here, in Figure 3(f), the inner yard is also reconstructed as a room with its ceiling height determined from the height of the surrounding ceilings.

normal vectors of the triangles on the reconstruction results of VoxIR. Table 2 presents an evaluation of reconstruction results on the four datasets for three buffer levels. The reported values refer to reconstruction results achieved when using a normal grid determined by the method from Section 3.1 from occupancy grids. In parentheses, the respective values are given, if instead of a mere occupancy grid, the normal grid derived directly from the triangle mesh is used. As can be seen, using discretized indoor mapping data without normal vectors only slightly diminishes the reconstruction quality of VoxIR on most datasets, with the exception of dataset ‘Residential House’ showing a decrease in completeness and correctness of around 10 %.

Note that the reported values for completeness reach values above 1. This is due to the fact, that VoxIR tends to reconstruct room surfaces with a thickness of several voxels while they are usually represented as surfaces with a thickness of one voxel in the discretized reference data. Thus, there are much more reconstructed voxels representing building geometry than reference voxels which can lead to there being more true positive voxels than reference voxels when using a buffer level $b > 1$.

4.3 VoxIR Applied to Point Clouds

Finally, the normal classification method presented in this work allows to evaluate the voxel-based indoor reconstruction VoxIR also on indoor mapping point clouds like those of the ISPRS benchmark on indoor modelling (Khoshelham et al. 2017). To this aim, the reference IFC models were converted to triangle meshes using IfcConvert 0.6.0. All triangle meshes belonging to building structures not represented in the point clouds (i.e. the outer surfaces of the volumetric geometries) were removed in a semi-manual process using Blender 2.9.4. Furthermore, the point clouds were manually cleaned as well, as they contain parts of the building environments that are not represented by the reference models.

The removed parts are marked in red in the occupancy grids of the discretized point clouds depicted in the first column of Figure 3. From the resulting occupancy grids, normal grids were determined with the method presented in Section 3.1. These were then used as input for VoxIR, with the resulting reconstruction voxel grids being depicted in the right column of Figure 3 (with voxels of semantic classes ‘Wall Opening’ and ‘Interior Object’ omitted as they are not considered in the evaluation). The reference triangle meshes manually derived from the IFC models were discretized as well and they are depicted in the middle column of Figure 3 to serve as reference data for quantitative evaluation. The results are presented in Table 3.
Figure 3. Visualization of input occupancy grids (manually removed parts depicted in red) on the left, discretized ground truth data in the middle and VoxIR reconstruction results on the right (both with 5 cm voxel resolution, parts of the ceilings are removed for better visibility) for the point clouds of the ISPRS benchmark on indoor modeling (Khoshelham et al., 2017, 2020). In the middle and right parts, ceiling is depicted in red, while floor and walls are depicted in green and grey, respectively.

Figure 4. Reconstruction results for two of the datasets depicted in Figure 3 when the red parts of the occupancy grids on the left in Figure 3 are included in the input occupancy grids.
As exemplified in Figure 4, the reconstruction results are also more or less reasonable when including the removed parts of the input occupancy grids depicted in red on the left-hand side of Figure 3, which are not considered in the reference IFC models. For instance, the lower floor of Figure 4(a) is partly reconstructed wherever there are parts of its floor or ceiling available in the input data. Also, the outdoor terrain around the building in Figure 4(b) is included in the reconstruction, with its height being determined by the height of the trees in the rearward part behind the building. In this case, the height of the ceilings reconstructed above the inner yards gets set to the tree height as well. Generally indoor/outdoor transitions and their consideration in modeling and automated reconstruction of buildings is still an interesting topic for future research [Previtali et al., 2014; Koch et al., 2016].

6. CONCLUSION

In this paper, we presented a novel method for normal classification of voxel occupancy grids of discretized indoor mapping data along with qualitative and quantitative evaluation results. Furthermore, the proposed method was used to extend an existing voxel-based indoor reconstruction pipeline to be applicable to input data in the form of point clouds. This enabled us to evaluate the indoor reconstruction approach, which so far was only applicable to triangle meshes, on the point clouds of the well-known ISPRS Benchmark for Indoor Modeling.

While the presented results are promising, there is still ample opportunity for further research left for future work. For instance, the evaluation presented here focuses exclusively on the reconstruction of building structures without regard for further semantics or room partitioning which are also considered in the evaluated voxel-based indoor reconstruction approach. Furthermore, the discussed method (for normal classification as well as for indoor reconstruction) is still restricted to the discretized voxel space. Conversion of the results towards actual surface or even volume geometries in Euclidean space would be a valuable extension, providing the means for the automated generation of actual BIM models from indoor mapping data and enabling better comparability of the results with those of other reconstruction methods. Still, we believe that voxels hold great potential for building-related analysis tasks [Gorte et al. 2019; Song et al. 2019; Wang et al. 2020].

References


